

第二节 数值变量资料的统计分析



统计描述

本节结构

- 一 数值变量资料的频数表
- 二 集中趋势
- 三 离散趋势
- 四 正态分布
- 五 抽样误差与参数估计
- 六 假设检验

数值变量资料的描述方法:

1、频数表与频数分布

2、统计指标

(1)、集中趋势指标：平均指标（算术均数、几何均数、中位数、众数、调和均数）

(2)、离散趋势指标：变异指标（极差、四分位间距、方差、标准差、变异系数）

3、统计表、统计图

— 数值资料的频数分布 frequency distribution

(一)、频数分布表

频数分布：指观察值在某组段出现的次数；频数表：为了解一组同质观察值的分布规律，在观察值个数（即样本含量， n ）较多时，可编制频数分布表，简称频数表。

- [例7-1]某校诊断学基础教研室为研究健康成年女性体温正常值，随机抽取102名健康(非排卵期)女大学生测试其体温，下列是测试午饭后休息一小时口腔温度($^{\circ}\text{C}$)的结果，试编制频数分布表。

例7-1 120名健康成年女性的口腔温度测定结果(°C)

编号	口腔温度	编号	口腔温度
1	37.0
2	36.9	61	<u>36.5</u>
3	37.2
4	37.1	97	36.6
5	37.0	98	36.9
6	36.8	99	36.9
7	36.8	100	36.9
...	...	101	36.7
41	<u>37.5</u>	102	36.8

1. 频数表的编制步骤

(1) 求全距 (range)：即最大值与最小值之差，又称为**极差**。

本例极差： $R=37.5-36.5=1.0(^{\circ}\text{C})$

(2) 确定**组距**和组限：根据研究目的和样本含量 n 确定。组距=极差/组数，通常分10-15个组，为方便计，组距参考极差的十分之一，再略加调整。

本例 $i = R / 10 = 1.0 / 10 = 0.1$ 。

列出组限，即组界限。每个组段的起点称下限，终点称上限。第一组段的**必须包含最小值**，最后一个组段**上限必须包含最大值**，其它组段上限值忽略。

(3) 列表归组：用划记法将所有数据归纳到各组段，得到各组段的频数。

第二节 数值变量资料的统计分析-一、二、

组段 (1)	划记 (2)	频数, f (3)	组中值, X (4)	fX (5)= (3)×(4)	fX^2 (6)= (3)×(4) ²
36.5~	—	1	36.55	36.55	1335.9025
36.6~	⊥	3	36.65	109.95	4029.6675
36.7~	正	5	36.75	183.75	6752.8125
36.8~	正正⊥	12	36.85	442.20	16295.0700
36.9~	正正正⊥	17	36.95	628.15	23210.1425
37.0~	正正正正正	25	37.05	926.25	34317.5625
37.1~	正正正一	16	37.15	594.40	22081.9600
37.2~	正正一	11	37.25	409.75	15263.1875
37.3~	正⊥	7	37.35	261.45	9765.1575
37.4~	⊥	4	37.45	149.80	5610.0100
37.5~37.6	—	1	37.55	37.55	1410.0025
合计	---	102		3779.80	140071.4750

(二)、频数分布图(直方图)

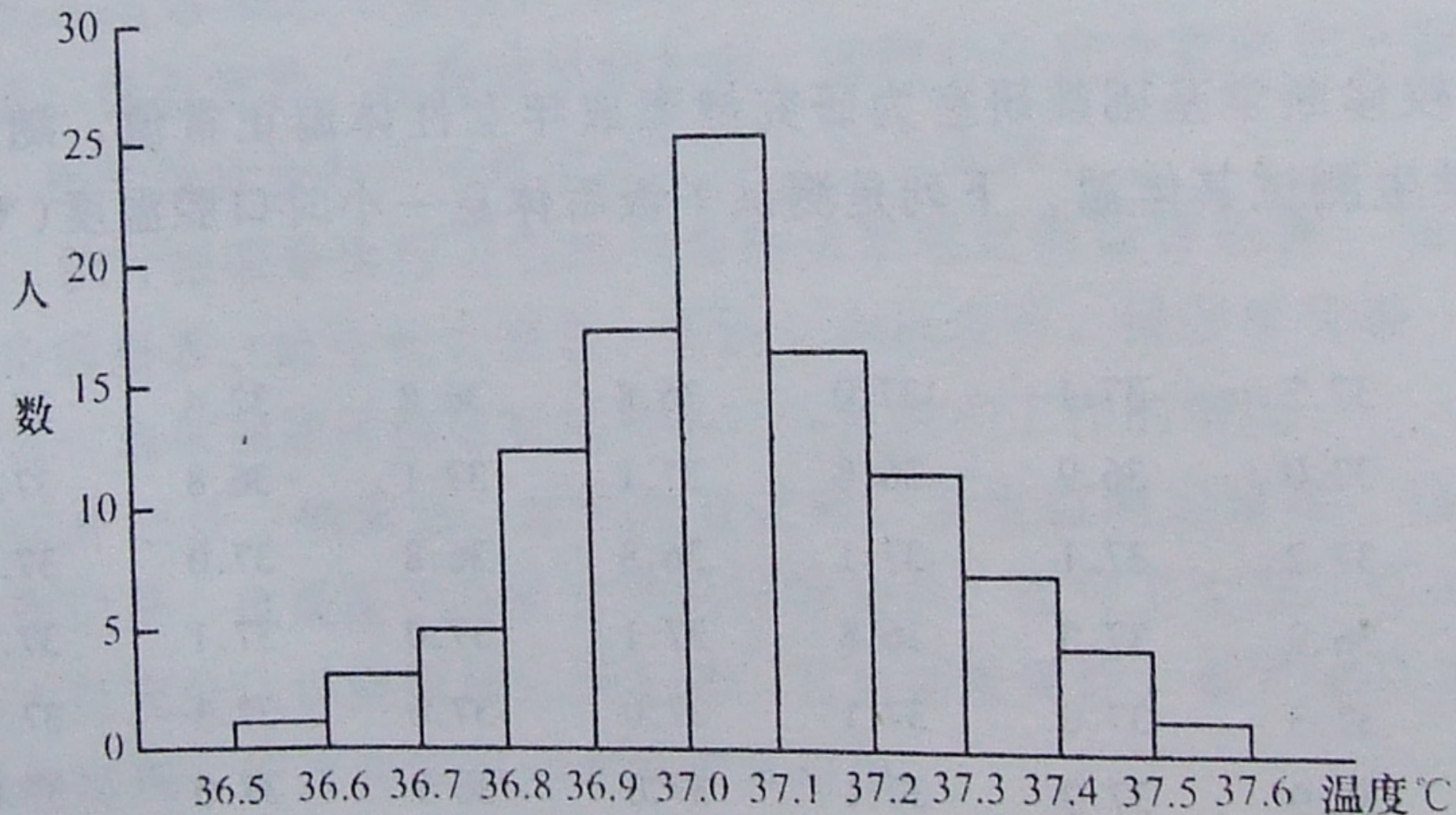


图 7-1 102名健康女大学生口腔温度频数分布

(三) 频数分布特征

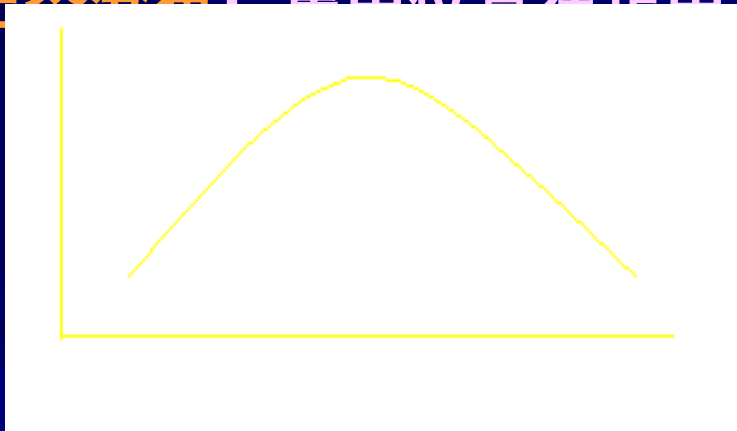
①集中趋势(central tendency): 变量值集中位置。本例在组段“37.0~”。**平均水平指标**

②离散趋势(tendency of dispersion): 变量值围绕集中位置的分布情况。离“中心”位置越远, 频数越小; 且围绕“中心”左右对称。**变异水平指标**

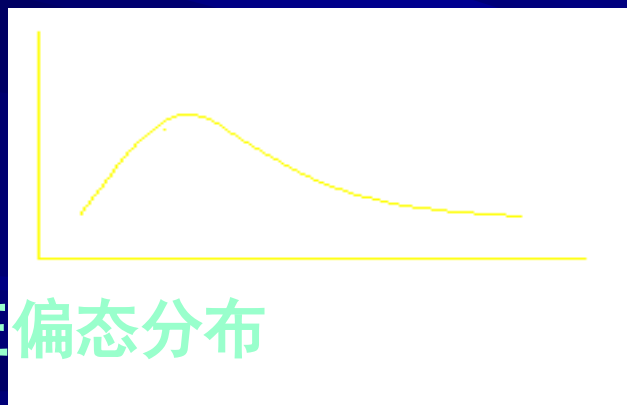
从不同角度说明被研究的事物。

(四) 频数分布类型

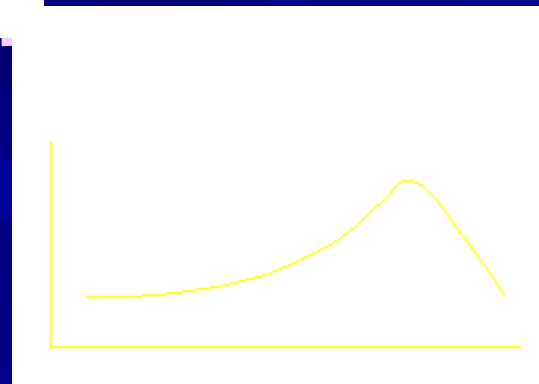
- ① 正态分布，集中位置在正中，左右两侧基本对称。



- ② 偏态分布：集中位置偏向一侧，左右两侧不对称。



正偏态分布



负偏态分布

分布类型不同，采用的统计方法不同。

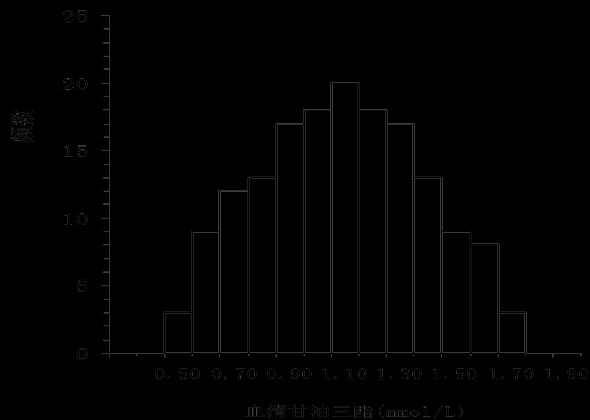
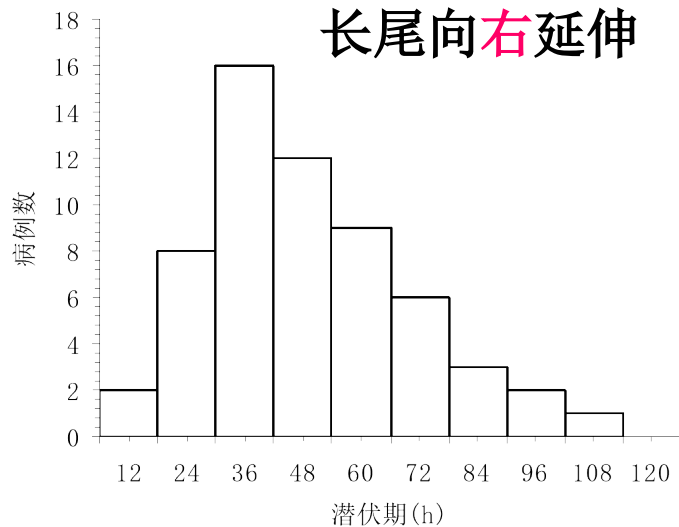


图12-1 160名正常成年女子的血清甘油三酯的频数分布图

态分布：中间高、
两边低、左右对称

正偏态分布：
长尾向**右**延伸



1

负偏态分布：
长尾向**左**延伸

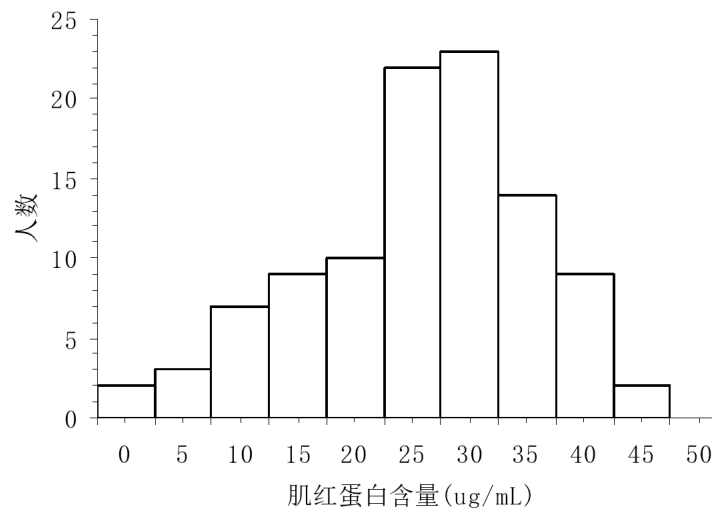


图2-3 101名正常人的血清肌红蛋白含量

第二节 数值变量资料的统计分析-

一、二、三

(五)、频数分布表的用途

- 1、揭示资料的分布类型；
- 2、显示频数分布的两个重要特征；
集中趋势 (Central tendency)
离散趋势 (Tendency of dispersion)
- 3、根据频数分布的不同类型，选择适当的统计方法，进行计算与分析；
- 4、利于发现某些特大或特小的可疑值。

二、集中趋势指标

1、算术均数

2、几何均数

3、中位数

平均数 (average)

描述一组性质相同的变量值的集中趋势或者集中水平的指标称为平均数，它是样本变量值或者总体变量值的代表值。根据资料的频数分布不同，可分别计算算术均数、几何均数和中位数。

(一)算术均数

(arithmetic mean)

算术均数是最常用的集中趋势指标，简称为均数（mean），是描述一组正态分布或者近似正态分布资料集中趋势的指标。样本均数以 \bar{x} 表示，总体均数以 μ 表示。

样本均数的计算方法

1. 小样本不分组资料 (直接法)

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{\Sigma X}{n}$$

2. 大样本分组资料 (加权法) (weighting method)

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \cdots + f_k X_k}{f_1 + f_2 + \cdots + f_k} = \frac{\Sigma fX}{\Sigma f}$$

均数的应用

- 1、均数反映一组同质观察值的平均水平，并可作为样本的代表值与其他样本进行比较。
- 2、均数适用于描述单峰对称分布，特别是正态分布或近似正态分布资料的集中趋势。
- 3、均数在描述正态分布特征方面具有重要意义。

(二)几何均数 (geometric mean)

当变量值的变化呈等比级数关系，特别是变量值的频数分布呈偏态分布，但经过对数转换后呈正态分布，即对数正态分布资料，适合于用几何均数描述其集中趋势，以符号G表示。

几何均数的计算方法

1. 小样本不分组资料 (直接法)

$$G = \sqrt[n]{X_1 X_2 \cdots X_n}$$

或

$$G = \lg^{-1} \left(\frac{\lg X_1 + \lg X_2 + \cdots + \lg X_n}{n} \right) = \lg^{-1} \left(\frac{\sum \lg X}{n} \right)$$

2. 大样本分组资料 (加权法)

$$G = \lg^{-1} \left(\frac{f_1 \lg X_1 + f_2 \lg X_2 + \cdots + f_k \lg X_k}{f_1 + f_2 + \cdots + f_k} \right) = \lg^{-1} \left(\frac{\sum f \lg X}{\sum f} \right)$$

应用几何均数注意事项

- 1、几何均数常用于等比级数资料，如抗体平均滴度和药物平均效价、卫生事业平均发展速度、人口的几何增长等，或用于对数正态分布资料；
- 2、观察值不能有0，因为0不能取对数，不能与任何其它数呈倍数关系；
- 3、观察值不能同时有正值和负值。若全是负值，计算时可把负号去掉，得出结果后再加上负号。
- 4、同一资料，几何均数 $<$ 均数。

血清的抗体效价滴度的**倒数**分别为：10、100、1000、10000、100000，求几何均数。

$$G = \lg^{-1} \left(\frac{\lg 10^1 + \lg 10^2 + \lg 10^3 + \lg 10^4 + \lg 10^5}{5} \right) = 1000$$

此例的算术均数为22222，显然不能代表滴度的平均水平。同一资料，**几何均数 < 均数**

(三) 中位数 (median)

把 n 个变量值由小到大顺序排列，位次居中的变量值称为中位数。适用于描述偏态分布资料 and 资料分布的末端无确切数据的开口资料的集中趋势，用符号 M 表示。中位数是一个特定的百分位数 P_{50} 。

中位数的计算方法

1. 小样本不分组

当n为偶数时:

$$M = \frac{1}{2} \left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)} \right)$$

当n为奇数时:

$$M = X_{\left(\frac{n+1}{2}\right)}$$

2. 大样本分组资料 (频数表法)

$$M = L + \frac{i}{f_m} \left(\frac{n}{2} - \Sigma f_L \right)$$

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/915023344002012010>